
Pomen in obseg oznak za jezikovne različice v okviru informacijske tehnologije

Han Steenwijk

Najprej so predstavljena različna merila in obsegi jezikovnih kod, ki so registrirane pri ISO 639-1, 639-2, 639-3 in 639-6. Potem sta predstavljena RFC 5646 in 4647, kar naj bi razložilo, kako se lahko oblikujejo pravilne oznake za jezikovne različice. Nazadnje je ponujen povzetek zadeve o vložitvi prošnje o kodi za rezijanščino.

Ključne besede: informacijska tehnologija, mednarodni standardi, kode za jezikovne različice, rezijansko narečje slovenščine

The meaning and scope of tags for linguistic varieties in information technology

This article first discusses the various criteria for and scopes of the language codes registered as ISO 639-1, 639-2, 639-3, and 639-6. It then introduces RFC 5646 and RFC 4647 in order to explain how well-formed language tags can be created. Finally, the case of applying for a language code for Resian is summarized.

Key words: information technology, international standards, language variety codes, Slovenian dialects, Resian dialect of Slovenian

1 Samostalniki, pojmi in krajšave

Občno ime in lastno ime se nanašata na pojme, ki se nanašajo na konkretne in abstraktne predmete v resničnem svetu. Tako jezik poimenuje te predmete. Odnosi med predmeti, naj bodo taksonomski, hierarhični ali kateri drugi, so vsebovani v teh pojmi. Na primer: poimenovanje *rezijanščina* se nanaša na jezikovno različico, ki jo srečujemo v resničnem svetu, in prav tako se poimenovanje *slovenščina* nanaša na jezikovno različico, ki jo srečujemo v resničnem svetu. Odnosi med tema dvema jezikovnima različicama niso neločljive lastnosti samih samostalnikov *rezijanščina* in *slovenščina*, ampak pojmov »rezijanščina« in »slovenščina«, na katera se ta dva samostalnika nanašata. Nič v samostalniku *rezijanščina* ne razkriva, da je lahko v hierarhičnem odnosu do samostalnika *slovenščina*. Samo pojma »rezijanščina« in »slovenščina« sta zmožna nedvoumno izkazati ta odnos.

Da bi se bolje seznanili s tema dvema pojmom, se je treba opreti na slovarje, pri lastnih in občnih imenih pa tudi na enciklopedije. Definicije v teh referenčnih delih so lahko semantično izčrpne ali ne, vendar vseeno vsebujejo določeno količi-

no informacij o pojmu. Za geslo *rezijanščina* ponuja Slovar slovenskega knjižnega jezika (SSKJ) razlago 'rezijansko narečje'. Tako torej izvemo, da se samostalnik nanaša na jezikovno različico, ki je v podrejenem hierarhičnem odnosu do neke druge, nedoločene jezikovne različice. Enciklopedija Slovenije (ES) vsebuje geslo *rezijansko narečje* in pojasnjuje, kje se to narečje govori in kako se je razvijalo v stoletjih od skupnega slovenskega fonološkega sistema, ki je osnova vseh sodobnih slovenskih narečij. Jezikovna različica, do katere je *rezijanščina* v podrejenem hierarhičnem odnosu, mora torej biti *slovenščina*.

Kar velja za samostalnike, velja tudi za krajšave teh samostalnikov. Krajšave se prav tako nanašajo na pojme, ki so lahko povezani med seboj ali pa tudi ne. Krajšava je uporabljena v jeziku v okoliščinah, ko ni nujno oz. ni zaželeno uporabljati celotnega imena, ker je pojem le-tega v danem kontekstu že dobro znan, ali pa zaradi časovne in prostorske gospodarnosti. Na splošno ne obstaja za vsak samostalnik ustrezna krajšava, obstajajo pa nadomestki za poimenovanja in imena, ki se v jeziku ali v določenih kontekstih pogosto pojavljajo. V razlagi termina *rezijansko narečje* v ES sta med drugim uporabljeni krajšavi *r. n.* in *slov.*, ki se nanašata na dva pojma, ki sta v danem kontekstu zelo dobro znana in se pojavljata dovolj pogosto.

2 Krajšave in pojmi v ISO 639

Širša javnost se je s krajšavami ISO za jezikovne različice seznanila z razmahom informacijske tehnologije. Kot je dobro znano, je učinkovita in univerzalna uporaba informacijske tehnologije močno odvisna od standardizacije. Na primer: arhitektura, na kateri sloni delovanje spleta, je bila ustvarjena na osnovi mnogih standardiziranih procesov. Standardizirane krajšave za jezikovne različice so ena izmed sestavin, ki jih informacijska tehnologija uporablja v svoji arhitekturi. Prav tako se tudi bibliografske zbirke podatkov, kakršne so postavile knjižnice, opirajo na standardizirane krajšave za opis kakega dokumenta. Te krajšave so registrirane v standardu ISO 639, ki je razdeljen na več delov, in sicer 639-1, 639-2 itd., ti pa se razlikujejo glede na zgodovinsko ozadje in merila za vključevanje kode za neko jezikovno različico, v dokumentaciji ISO imenovano »jezik«, na svoj seznam. Razlika med kodami in krajšavami je, da ni treba, da koda odseva samostalnik oz. pridevnik za ime jezikovne različice v njeni formalni obliki. Ta formalna ustreznost ne more biti zagotovljena v vseh primerih. Izmed vseh delov ISO 639 so merila, ki so uporabljena v ISO 639-1 »Codes for the representation of names of languages, Part 1: Alpha-2 code«, najbližja temu, kar bi pričakovali za izbiro jezikovne različice na vrhu neke sociolingvistične hierarhije, tj. polno delujočega knjižnega jezika:

- (1) velik nabor obstoječih dokumentov (specializiranih besedil, kot so srednješolski ali univerzitetni učbeniki, priročniki za tehnično dokumentacijo, strokovne revije, knjige za različna področja itd.), napisanih v strokovnem jeziku;
- (2) določeno število obstoječih terminologij raznovrstnih področij (kot so tehnični slovarji, specializirani glosarji, slovarji itd. v tiskani ali elektronski obliki);

- (3) priporočilo in podpora strokovne ustanove (kot je organizacija za standardizacijo, vladno telo, jezikoslovna ali kulturna ustanova).

[ISO 639-2: Criteria for ISO 639-1]

Nadaljnje odločilne točke so:

- (1) število govorcev jezikovne skupnosti;
- (2) priznan status jezika v eni ali več državah;
- (3) podpora prošnji s strani enega oz. več uradnih teles.

[ISO 639-2: Criteria for ISO 639-1]

Tako med dvočrkovnimi kodami, registriranimi v ISO 639-1 [ISO 639-2: Language Code List], najdemo krajšave za splošno znane knjižne jezike, in to žive in mrtve. Kljub temu pa so na seznam uvrščene tudi kode za manj znane evropske knjižne jezike, kot so bretonščina, retoromanščina in zahodna frizijščina. Glede na vprašanje hierarhičnega statusa določene jezikovne različice pa ISO 639-1 nima izrecnih meril: »Splošno gledano narečja ne dobijo kod, vendar pa je razlikovanje med narečji in jeziki določeno v vsakem primeru posebej« [ISO 639-2: Frequently asked questions]. Na vključitev limburščine, jezikovne različice, ki se govori v Belgiji, na Nizozemskem in v Nemčiji, je gotovo vplivalo dejstvo, da je na Nizozemskem uradno priznana kot pokrajinski jezik. Po drugi strani pa bretonščina nima takega statusa, vendar je na srečo jezikovno strogo ločena od francoščine, državnega jezika tistega dela Evrope.

ISO 639-2 »Codes for the representation of names of languages, Part 2: Alpha-3 code« se osredotoča na tričrkovne kode za pisne ali kako drugače dokumentirane jezikovne različice, za katere pa ni nujno, da so polno delujoči knjižni jeziki. Merila za izbiro so:

- (1) Število dokumentov. Prošnja za jezikovni identifikator bo vključevala dokaz, da ima ena agencija 50 različnih dokumentov v tem jeziku oz. da ima pet agencij skupaj 50 različnih dokumentov v tem jeziku. Dokumenti so vse oblike gradiva, ne le pisne. To je nujna zahteva, vendar sama po sebi ne zadošča. Zato se bodo upoštevale še druge zahteve.
- (2) Obseg in raznovrstnost literature v tem jeziku, najsi bo v pisni ali ustni obliki, bosta upoštevana in bi morala biti v predlogu dokumentirana. Dokumentacija je lahko v obliki knjižničnih referenc ali bibliografije ali splošnejših izjav, ki naštevajo literaturo in njeno raznovrstnost.
- (3) Državna ali pokrajinska podpora. Zaželeno je, da je predlog izrecno podprt s strani državne ali pokrajinske avtoritete oz. telesa, ki standardizira. Če take podpore zaradi katerega koli razloga ni, se bo upoštevalo priporočilo kake druge avtoritete ali jezikovne organizacije.
- (4) Formalni ali uradni status. Če ima obravnavani jezik kakršen koli uradni status, dokumentiranost tega statusa močno podpira predlog. Dodelitev formalnega statusa jezikov nikakor ni nekaj, kar bi bilo po svetu konsistentno izvajano, in pomanjkanje le-tega ni negativen argument, če so izpolnjene preostale zahteve.

(5) Formalna izobrazba. Če je jezik uporabljen kot sredstvo poučevanja v izobraževanju na kateri koli ravni, bo dokumentiranost o tej rabi podprla predlog. Učenje jezika je prav tako pomembno, še posebno, če je učenje množično.

[ISO 639-2: Criteria for ISO 639-2]

»Neke vrste uradni status« ima tu velik pomen, vendar ne odločilnega. Najpomembnejša je zahteva po dovolj velikem številu dokumentov v določeni jezikovni različici. Število jezikovnih različic, vključenih v seznam 639-2 [ISO 639-2: Language Code List], je veliko večje kot tistih v ISO 639-1. Po definiciji so vse jezikovne različice, vključene v ISO 639-1, vključene tudi v 639-2. Med manj znanimi knjižnimi evropskimi jeziki so nekateri uvrščeni samo v 639-2, npr. furlanščina, kašubščina, zgornja in spodnja lužiščina, nizka nemščina, neapeljščina, severna in vzhodna frizijščina, romščina, sicilijanščina in škotščina. ISO 639-2 se pri perečem vprašanju »narečje ali jezik« ponovno ravna po pravilih, ki jih določa ISO 639-1: »O razlikah med narečji in jeziki se bo odločalo v vsakem primeru posebej« [ISO 639-2: Criteria for ISO 639-2]. Nekatere enote z zgornjega seznama evropskih jezikov, kot sta kašubščina in sicilijanščina, očitno niso polno delujoči (polnofunkcionalni) knjižni jeziki. Še več: zadnji nima nikakršnega formalnega statusa.

Niti v ISO 639-1 niti v ISO 639-2 eksplicitni zapisi z informacijo o posameznih kodah niso na razpolago. Pojmi, na katere se kode nanašajo, ostajajo zunaj teh delov ISO 639. ISO 639-1 in 639-2 ne »registrirata jezikov«, ampak le kode za imena jezikov. Čeprav je načrt teh delov ISO zelo premišljen, pa izpust podatkovnih zapisov pušča prostor za ugibanja o merilih za vključitev ali izključitev določene jezikovne različice. Zakaj sta na primer retoromanščina in zahodna frizijščina sestavini ISO 639-1, furlanščina ter severna in vzhodna frizijščina pa »le« sestavini ISO 639-2? Zakaj pa beneščine ni na nobenem od seznamov, čeprav neapeljščina in sicilijanščina v ISO 639-2 sta? Morda le zato, ker za kodno sestavino za beneščino ni bila nikoli podana prošnja, ali pa zaradi česa drugega?

V nekaj pogledih se ISO 639-3 »Codes for the representation of names of languages, Part 3: Alpha-3 code for comprehensive coverage of languages« v pozitivnem smislu razlikuje od do zdaj obravnavanih delov ISO 639. Prvič: meri na to, da bi izdelal in ohranil obširen seznam vseh človeku znanih jezikov, ne pa le izbora. Drugič: vsaka kodna sestavina ima podatkovni zapis, ki je povezan z njo. Torej sta koda in pojem v tem delu ISO 639 v medsebojnem odnosu. Tretjič: v primerjavi z ISO 639-1 in 639-2 ima jasnejši pristop do vprašanja »narečje ali jezik«:

Ni niti ene definicije termina jezik, s katero bi se strinjali vsi in ki bi bila primerna za vse namene. Zato lahko pride do nestrinjanja celo med govorce ali jezikoslovci o tem, ali dve različici predstavljata narečji enega jezika ali sta to dva različna jezika. Ta del standarda ISO o tem, ali sta dve različici del istega jezika ali sta dva različna jezika, razsoja na podlagi več dejavnikov, med katerimi so jezikovna podobnost, razumljivost, skupna literatura, pogled govorcev na odnos med jezikom in identiteto idr. Osnovna merila so tale:

(1) dve podobni različici sta po navadi ocenjeni za različici istega jezika, če so govorce ene različice sposobni razumevanja druge (tj., če lahko z znanjem

svoje različice razumejo drugo različico, ne da bi se je morali učiti) na funkcionalni ravni;

(2) kjer je razumevanje med govorniki različic omejeno, je lahko obstoj skupne literature ali skupne etnojezikovne identitete s središčno različico, ki jo razumejo govorniki obeh različic, močan pokazatelj, da morata biti kljub vsemu obravnavani kot različici istega jezika;

(3) kjer je dovolj razumevanja med uporabniki obeh različic, da je omogočeno sporazumevanje, je lahko obstoj dobro osnovane ločene etnojezikovne identitete močan pokazatelj, da morata biti kljub vsemu obravnavani kot dva različna jezika.

[ISO 639-3: Scope of denotation for language identifiers]

Termin *narečje* pa je tudi konkretiziran:

Za jezikovne različice, označene z vsakim izmed identifikatorjev v tem delu ISO 639, se predvideva, da so ločeni jeziki in ne le narečja drugega jezika, čeprav za nekatere namene uporabniki lahko domnevajo, da je različica, ki je na seznamu v tem delu ISO 639, prej narečje kot jezik. Po tem standardu je termin *narečje* uporabljen kot na področju jezikoslovja, kjer preprosto identificira vsako podrazličico jezika, ki bi lahko obstajala glede na geografsko pokrajino, starost, spol, socialni razred, obdobje ipd. To se razlikuje od nestrokovne rabe, v kateri je narečje tipično razumljeno tako, da bi konotiralo podstandard ali manj razvite oblike jezika.

Narečja nekega jezika so vključena v denotacijo označbe, ki jo predstavlja identifikator jezika. Vsak jezikovni identifikator torej predstavlja popoln obseg vseh govornih in pisnih različic tega jezika, vključene pa so tudi vse standardizirane oblike.

[ISO 639-3: Scope of denotation for language identifiers]

Zato ni nikakršno presenečenje, da se v ISO 639-3 nahajajo jezikovne različice z nizkim hierarhičnim položajem, kot so kölnščina (*kölsch*), bavarščina, alemanščina, vsa nemška pokrajinska narečja brez vsakršnega uradnega statusa, čeprav so pogosto rabljena v mnogih funkcijah, npr. v gledališču in na televiziji.

Podatkovni zapis, povezan z vsako od kodnih sestavin, vsebuje tele attribute: prebivalstvo (število govorcev), pokrajino, variantna imena, narečja, klasifikacijo, rabo jezika, razvoj jezika in pisni sistem. Vse skupaj dopuščajo možnost za dobro definiran pojem. Atribut »klasifikacija« lahko vsebuje vso genetsko informacijo o tej jezikovni različici, atribut »raba jezika« pa nas informira o funkcionalnosti.

Četrtrič, ISO 639-3 na vrhu hierarhije bolj eksplicitno uporabi pojem »makrojezik«, ki pokriva zelo različna narečja ali pa mrežo zelo sorodnih jezikov. Na primer: koda *hbs* v ISO 639-3 označuje makrojezik 'srbohhrvaščina', na katero so navezane hrvaščina, srbsščina in bosanščina. Prav tako ima makrojezik »albanščina« (*sqi*) nase navezane narečne skupine albanščino *arbëreshë*, albanščino *arvanitika*, gegovsko albanščino in toskovsko albanščino, makrojezik »sardščina« (*srd*) pa narečne skupine *campidanese*, *gallurese*, *logudorese* in *sassarese*. Vsi ti individualni jeziki in narečne skupine imajo v ISO 639-3 pripisane ločene kodne sestavine.

Od 17. novembra 2009 obstaja tudi ISO 639-6 »Codes for the representation of names of languages, Part 6: Alph-4 code for comprehensive coverage of language variants«. Ta del se v marsičem razlikuje od doslej obravnavanih delov ISO 639. Prvič so vse registrirane jezikovne različice izrecno postavljene v medsebojen hierarhični odnos. Na primer:

```
aeng eng Anglo-English
[...]
seng aeng Anglo-English Spoken
[...]
cang seng Anglo-English Central Cluster
aenc cang Anglo-English West Central Cluster
scou aenc Scouse
[ISO FDIS 639-6 en-GB]
```

Tukaj je kodna sestavina *scou* iz ISO 639-6 postavljena v hierarhični odnos do kodne sestavine *eng* iz ISO 639-2, in sicer v smislu verige starši–otrok: *eng* → *aeng* → *seng* → *cang* → *aenc* → *scou*. Poleg tega da vključuje jezikovne različice z nizko hierarhično pozicijo, tj. narečja, so lahko v štiričrkovne kode vključeni tudi drugi parametri jezikovne variacije, kot so »govorjeno ali pisno«, uporabljena pisava in »zgodovinsko ali sodobno«.

Kar se tiče rezijanščine, so v ISO 639-6 predlagane štiri kode, in sicer za »rezijanščino«, »pisno rezijanščino«, »pisno rezijanščino, pisano v latinici« in »govorjeno rezijanščino«. Čeprav imajo vsa večja slovenska narečja, kot so prekmursko ali ziljsko, predvidene kodne sestavine, pa nima nobeno drugo, razen rezijanščine, razlikovalnega kodne sestavine za govorjeno ali pisno različico.

3 Uporaba in procesiranje jezikovnih kod ISO 639

Naloga ISO 639 je bila in je, da določi, katere kode obstajajo za identificiranje jezikovnih različic, vendar pa ni nič rečeno, kako naj bi se te kode uporabljale. Na tem mestu si lahko pomagamo z RFC 5646 »Tags for Identifying Languages« in RFC 4647 »Matching of Language Tags«. RFC 5646 na primer priporoča rabo dvočrkovnih kod, kjer obstajata obe, dvo- in tričrkovna. Ali manj trivialno: ukvarja se z načini, kako kodno sestavino ISO 639 razširiti s povezovanjem z drugimi kodami, tako da dobimo pravilen niz oznak, ki omogoča podrobnejši opis informacijskega predmeta. Te kode izhajajo iz drugih standardov ISO, kot sta ISO 3166-1 »Codes for the representation of names of countries and their subdivisions, Part 1: Country codes« in ISO 15924 »Codes for representation of names of scripts«, ali iz registra IANA za variantne podoznake [IANA Language Subtag Registry]. Primera takih nizov oznak sta *sr-Latn-ME*, kar pomeni 'srbsščina, pisana v latinici, kot se uporablja v Črni gori', in *de-CH-1996* 'nemščina, kot se uporablja v Švici, po pravopisu iz leta 1996'. V teh nizih oznak sta *de* in *sr* kodi ISO 639-1, *Latn* je koda ISO 15924, *CH* in *ME* sta kodi ISO 3166-1 in *1996* je variantna podoznaka, registrirana pri

IANA. V okolju, v katerem je ustrezen le del niza oznak, je procesiranje niza odvisno od hierarhičnih odnosov, ki veljajo med podoznakami. Na primer: v računalniškem programu, kjer je za vse vrste nemščine na razpolago le *de*, bo vnos oznak, kot sta *de-Latf* 'nemščina, pisana v fraktorni različici latinice' in *de-CH-1996*, dal kot rezultat informacijski objekt z oznako *de*, ker se domneva, da imajo tako bralci frakture kot nemški govorniki iz Švice osnovno znanje nemščine nasploh.

Na prvi pogled se zdi, da lahko pokrajinske kode ISO 3166 dajo nekatere možnosti za oblikovanje pravega niza oznak za zemljepisna narečja. Vendar pa je bolj stvar naključja, da se tako osnovana vrstica pokriva z nekim bolj ali manj definiranim geografskim narečjem. Znotraj konteksta slovenščine se npr. *sl-HU* 'slovenščina, kot se uporablja na Madžarskem' za porabščino zdi smiselno, vendar je narečno realnost za *sl-AT* 'slovenščina, kot se uporablja v Avstriji' in *sl-IT* 'slovenščina, kot se uporablja v Italiji' veliko težje definirati. Treba je opozoriti na to, da koda *sl* zajema slovenščino v vsej njeni hierarhični variantnosti in torej interpretiranje *sl-HU* kot 'slovensko narečje, kot se govori na Madžarskem' ni trivialno, ni pa izključeno. Če se ukvarjamo z manjšimi geografskimi enotami, ki so predstavljene v ISO 3166-2 »Codes for the representation of names of countries and their subdivisions, Part 2: Country subdivision codes«, se povečuje možnost sovpadanja med nizom oznak in zemljepisnim narečjem, vendar samo to. V Italiji razlikujejo dežele in pokrajine, v Sloveniji pa trenutno samo občine. Poleg tega pa RFC 5646 ne predvideva, da so kodne sestavine ISO 3166-2 uporabljene kot podoznake.

Veliko primernejše za razlikovanje zemljepisnih in drugih narečij znotraj okvira RFC 5646 so variantne podoznake: »Variantne podoznake so uporabljene kot pokazatelji za dodatne dobro prepoznavne različice, ki definirajo neki jezik ali njegova narečja, ki niso označena s kako drugo razpoložljivo podoznako« [RFC 5646, razdelek 2.2.5] in »Narečne ali katere druge razlike ali različice znotraj jezika, njegov pravopis, pisni sistem, pokrajinsko ali zgodovinsko rabo, prečrkovanje ali druge spremembe ali razlikovalne variacije so lahko registrirani kot variantne podoznake« [RFC 5646, razdelek 3.6]. Podoznake, ki so bile registrirane do zdaj, se nanašajo npr. na pravopis (*1901*, *1996* za rabo z *de*), pisni sistem (*fonipa* 'mednarodna fonetična abeceda', tj. 'International Phonetic Alphabet'), zgodovinsko rabo (*1694acad* za rabo s *fr*) in pokrajinska narečja (*scouse* za rabo z *en*).

RFC 5646 prav tako predvideva rabo določenih kodnih sestavin ISO 639-3 kot t. i. razširjenih jezikovnih podoznak skupaj s kodami za makrojezik. Kantonska kitajščina je lahko označena z dvema – *zh-yue* ali preprosto *yue*, kjer je koda ISO 639-1 *zh* za kitajščino in koda ISO 639-3 *yue* za jezik *yue*, imenovan tudi kantonska kitajščina. Priporoča pa se, da se uporablja krajša oblika oznake (RFC 5646, razdelek 2.2.2). Trenutno ni nobenih jasnih predlogov, kako bi se kodne sestavine ISO 639-6 uporabile, da bi ustvarili pravilne nize oznak. Samo ugibamo lahko, da jih bo nekega dne mogoče uporabiti kot prve kodne sestavine v nizih oznak, saj so štiričrkovne jezikovne kode na tem mestu rezervirane za rabo v prihodnosti (RFC 5646, razdelek 2.2.1).

4 Rezijanščina in ISO 639-3

Med raziskovanjem rezijanščine je avtor začutil potrebo po oznakah, da bi identificiral ne le rezijanščino kot celoto, ampak tudi vse njene različice. Zato je zaprosil za variantne podoznake pod RFC 5646 ali, bolje, pod enim izmed njegovih predhodnikov, RCF 3066. Vendar pa je pri registriranju podoznake *rozaj* za rezijanščino oktobra 2003 diskusijska skupina za jezikovno označevanje IETF menila, da za rezijanščino z vsemi njenimi atributi (lokalna narečja, pisni jezik, zgodovina dokumentiranja) pridejo nadaljnje prošnje za registracijo variantnih podoznakov v pošteve šele po registraciji kode za rezijanščino kot tako pri ISO 639. V tistem času sta obstajala le ISO 639-1 in 639-2, zato je bila podana prošnja pri ISO 639-2. Na to ni bilo nikakršnega odziva, ne pozitivnega ne negativnega, in avtorju je bilo svetovano, naj poskusi srečo pri ISO 639-3, ki je bil tik pred tem, da ga uradno priznajo kot del ISO 639. Izmed vseh delov ISO 639 je ta najprimernejši, saj je v osnovi popoln in skrbi za podatkovne zapise. Še enkrat so poudarjene pravkar navedene značilnosti, saj se le s pomočjo podatkovnega zapisa lahko razloži hierarhično mesto rezijanščine glede na slovenščino kar se da eksplicitno.

Zato je bil junija 2006 poslan osnutek prošnje za vložitev pri ISO 639-3 skupini strokovnjakov za rezijanščino, ki delujejo v Estoniji, Italiji in Sloveniji. V tem obdobju je bilo predlaganih nekaj popravkov osnutka in besedilo je bilo revidirano v drugo verzijo. Nekako pa je predlog, verjetno njegova nerevidirana oblika, postal znan širši slovenski javnosti in kmalu zatem je Slovenska akademija znanosti in umetnosti v Ljubljani izrazila željo po diskusiji o predlogu na srečanju, predlaganem za september 2006, ki pa je bilo potem šele decembra istega leta. Poleg treh strokovnjakov iz Italije so bili vsi drugi udeleženci slovenski jezikoslovci in dialektologi. Po razjasnitvi nekaterih nesporazumov o precej idiosinkretični terminologiji obrazca prošnje ISO 639-3, še posebno pomena, pripisanega leksemu *language*, so se udeleženci diskusije odločili, da nadaljujejo s prošnjo, in se strinjali z na novo revidiranim besedilom, ki je vodilo v tretjo verzijo.

S pomočjo prvega obrazca [MS 1] je bilo zaprošeno za spremembo v pomenu v že obstoječi kodni sestavini, tj. v našem primeru *slv* 'slovenščina'. Taka sprememba v pomenu kodne sestavine je v terminologiji ISO 639-3 imenovana razcep. V tem obrazcu morajo biti navedeni razlogi za prošnjo za novo kodo za različico, ki jo je prej pokrivala koda *slv*. Na tem mestu je torej govor o merilih, s katerimi ISO 639-3 ločuje med jeziki in narečji. Tretja verzija besedila, o kateri so se prisotni strinjali, se glasi:

3. Razcepiti jezikovno kodo na dve ali več kodnih sestavin

(a) Seznam jezikov, na katere naj bo kodna sestavina razcepljena:

slovenščina, rezijanščina

[...]

(b) Po zgornjih merilih naj bo dana racionalna utemeljitev za razcepitev obstoječe kodne sestavine na dva ali več jezikov:

1. Knjižna slovenščina in rezijanščina med seboj nista razumljivi brez predhodnega učenja, vendar pa obstaja malo sporazumevalnih težav z govorcami sosednjih slovenskih narečij. Govorcu slovenščine kot prvega jezika se bo rezijanščina zdela zelo težko razumljiva, vendar pa bodo govorce rezijanščine kot prvega jezika lahko razumeli nekaj osnovne slovenščine.
 2. Literaturo, tiskano v slovenščini, razume le malo govorcev rezijanščine in je ne dojemajo kot del svoje etnolingvistične identitete.
 3. Čeprav imajo govorce slovenščine rezijanščino za slovensko narečje, se večina Rezijanov s tem pogledom ne strinja.
- [...]

5. Viri informacij

[...]

(c) *Védenje iz tiskanih virov (prosimo za popolne bibliografske podatke)*

Roberto Dapit, »Nastajanje krajevnih knjižnih jezikov pri Slovencih v Furlaniji«, v: *Slovenski knjižni jezik: aktualna vprašanja in zgodovinske izkušnje*, ur. A. Vidovič Muha, Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, 2003 (Obdobja 20), str. 301–312.

Aleksandr D. Duličenko, »Das Resianische«, v: *Einführung in die slavischen Sprachen*, 3. izdaja, ur. P. Rehder, Darmstadt: Wissenschaftliche Buchgesellschaft, 1998, str. 246–249.

Tine Logar, »Solbica (Stolvizza; OLA 1)«, v: *Fonološki opisi srpskohrvatskih/hrvatskosrpskih, slovenačkih i makedonskih govora obuhvaćenih opšteslovenskim lingvističkim atlasom*, Sarajevo, 1981, str. 35–40.

Pavle Merku, »Enkratnost Rezije: navezanost na kulturne korenine in shizoglosija«, *Jezik in slovstvo* 31 (1985/86), št. 4, str. 101–107.

Fran Ramovš, »The Resianic dialect«, in: *The National Frontier Between Slovenia and Italy: Linguistic and Social Features*, Ljubljana: J. Blaznika nasl., d. d., 1945, str. 5–7.

Matej Šekli, »Jezik, knjižni jezik, pokrajinski oz. krajevni knjižni jezik: genetskojezikoslovni in družbenostojezikoslovni pristop k členjenju jezikovne stvarnosti (na primeru slovenščine)«, v: *Aktualizacija jezikovnozvrstne teorije na Slovenskem: členitev jezikovne resničnosti = Obdobja 22*, ur. E. Kržišnik, Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, 2004, str. 41–58.

Han Steenwijk, »Resian as a minority language«, v: *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches*, ur. M. Janse – S. Tol, Amsterdam – Philadelphia: John Benjamins, 2003 (Current Issues in Linguistic Theory 240), str. 215–226.

Irena Zuljan, »Rezija in rezijanščina sta res enkratni: sociolingvistični izsledki jezikoslovne skupine mladinskega raziskovalnega tabora Rezija 1989«, *Primorska srečanja* 1989, št. 102, str. 913–918.

Lew N. Zybatow, »Resianisch: eine bedrohte Mikrosprache im Herzen Europas«, v: *Neue Forschungsarbeiten zur Kontaktlinguistik*, ur. W. W. Moelleken – P. J. Weber, Bonn: Dümmler, 1997 (Plurilingua 19), str. 505–514.

[MS 1]

Vsem udeležencem se je tak opis resnične sociolingvistične situacije zdel pošten in s pomočjo citirane literature, ki vsebuje tudi nasprotno poglede na problematiko,

naj bi bilo urednikom ISO 639-3 prepuščeno, da odločijo o statusu rezijanščine. Avtor zdaj za nazaj meni, da bi bilo nujno kaj povedati tudi o tradicionalni ustni književnosti, ki si jo delita rezijanščina in preostalo slovensko jezikovno področje.

V spremljevalnem drugem obrazcu [MS 2] je bilo zaprošeno za novo kodno sestavino za rezijanščino, dodanih pa je bilo kar največ podatkov, ki bi služili za podatkovni zapis. Poglavitni deli tega obrazca v njegovi tretji verziji so:

3. Način rabe in jezikovna pripadnost

(a) *Ta jezik je:* govorjen.

(b) *Jezikovna družina, če je klasificirana; izvor, če je nenaravno skonstruiran:* indoevropsčina, slovanščina, južnoslovanščina, zahodna južnoslovanščina, slovenščina, severozahodna slovenščina.

(c) *Jezikoslovno najbližji jezik. Za znakovne jezike omenite vpliv drugih znakovnih ali govornih jezikov:* severozahodna slovenska narečja.

4. Razvoj in raba jezika

(a) *Katera pisna literatura, zapisi ali poročila obstajajo v tem jeziku? Ali obstajajo časopisi, radijske ali televizijske oddaje itd.?*

Rezijanska zapisana besedila nastajajo od 18. stoletja dalje. Vse do sedemdesetih let prejšnjega stoletja so bila ta besedila pretežno verskega značaja. Dve rezijanski reviji – *All’Ombra del Canin* in *Näs glas* – redno objavljata besedila v jeziku kraja. Vsako soboto radijska postaja RAI predvaja 30-minutno oddajo v rezijanščini.

(b) *Je ta jezik uradno priznan s strani katere koli vladne veje? Je uporabljen kot jezik učenja ali poučevanja na katerikoli ravni formalnega izobraževanja za druge predmete?*

Rezijanščina ni uradno priznana in se v izobraževanju ne uporablja. Kljub temu pa se jo da učiti v osnovni šoli.

[MS 2]

S temi značilnostmi je jasno, da je rezijanščina eno izmed slovenskih narečij in da pokriva le omejen nabor funkcij. Na istem srečanju na akademiji se je razpravljalo tudi o možnostih za vložitev drugih podobnih prošenj za druge slovenske pokrajinske pisne jezike in v ta namen je bila ustanovljena komisija. Besedilo za prošnjo za kodo za rezijanščino, ki je bilo revidirano med zasedanjem, je avtor pripravil in ga poslal akademiji, da bi mu ta dala svojo zadnjo odobritev.

Od decembra 2006 do februarja 2007 je bilo v zasebni korespondenci med predsedujočim decembrskega srečanja in avtorjem načetih nekaj kočljivih točk v tretji verziji. Predsedujoči je poudaril tri take točke: (1) izključna raba termina *jezik* v besedilu obrazcev; (2) odnos rezijanščine do sosednjih slovenskih narečij; (3) nevarnost, da bi bila rezijanščina predstavljena kot jezikovna različica, ločena od slovenščine. Avtorjevo stališče do teh točk je bilo: (1) dano besedilo obrazcev je nespremenljivo in termin *jezik* mora biti sprejet v pomenu, ki ga določa ISO 639-3; (2) razdelek 3.b.1 prvega obrazca se lahko razširi in odnos rezijanščine do sosednjih slovenskih narečij se lahko razloži kot: »Knjižna slovenščina in rezijanščina med seboj nista razumljivi brez predhodnega učenja, vendar pa

obstaja malo sporazumevalnih težav z govorcami sosednjih slovenskih narečij [npr. z govorcami terskega narečja, ki živijo takoj južno od Rezije, in z govorcami obsoškega narečja, ki živijo neposredno vzhodno od Rezije]. Govorcu slovenščine kot prvega jezika se bo rezijanščina zdela zelo težko razumljiva, vendar pa bodo govorce rezijanščine kot prvega jezika lahko razumeli nekaj osnovne slovenščine.◀; (3) razdelek 3.b drugega obrazca popolnoma jasno kaže, da je rezijanščina del slovenskega jezikovnega področja.

Navsezadnje je bil maja 2007 avtor seznanjen s tem, da akademija ni sprejela revidiranega osnutka prošnje, saj se ji je zdelo, da je angleški leksem *language*, ki se nekajkrat pojavlja v obrazcu, nepremostljiva ovira. Zato se je avtor odločil, da prošnje ne preda ISO 639-3, saj po njegovem mnenju standardi lahko delujejo le, če so za uporabnika nesporni. Kaj se je dogajalo s komisijo za podobne prošnje, ustanovljeno na srečanju decembra 2006, avtorju ni znano.

5 Rezijanščina in RFC 5646

Z zgoraj opisanimi rezultati se je avtor ponovno obrnil na diskusijsko skupino za jezikovno označevanje IETF. Brez nadaljnjega so bile julija 2007 registrirane tele podoznake v RFC 4646, ki je neposredni predhodnik RFC 5646:

Subtag: 1994. Description: Standardized Resian orthography

Subtag: biske. Description: The San Giorgio dialect of Resian. Description: The Bila dialect of Resian

Subtag: lipaw. Description: The Lipovaz dialect of Resian. Description: The Lipovec dialect of Resian

Subtag: njiva. Description: The Gniva dialect of Resian. Description: The Njiva dialect of Resian

Subtag: osojs. Description: The Oseacco dialect of Resian. Description: The Osojane dialect of Resian

Subtag: solba. Description: The Stolvizza dialect of Resian. Description: The Solbica dialect of Resian

[IANA Language Subtag Registry]

Zdaj so lahko oblikovani nizi oznak, kot so *sl-rozaj-1994*, *sl-rozaj-biske*, *sl-rozaj-biske-1994* itd. Ti nizi oznak se uporabljajo npr. v bibliografskih zapiskih po standardu DCMI, ki opisujejo rezijanska besedila, objavljena na spletni strani *Resianica*. Kljub temu da ti nizi oznak zadovoljujejo najbolj osnovne označitvene potrebe in povsem jasno nakazujejo hierarhijo, taka rešitev ostaja nezadovoljiva s kar nekaj vidikov. Prvič so nizi oznak precej dolgi in okorni za vsakdanjo rabo in drugič, kar je še pomembneje, ujemalni algoritem za nize oznak vrne informacijske predmete, ki za uporabnika niso primerni. V brskalniku se lahko uporabnik na primer odloči za svoj(e) priljubljen(e) jezik(e) za obiskane strani. Domnevajmo, da je govorec rezijanščine specificiral *1. rezijanščina* in *2. italijanščina*, kar je potem enkodirano kot *1. sl-rozaj* in *2. it*. Ker ne obstaja veliko strani, ki bi ponujale rezijanski HTML, bo brskalnik uporabil ujemalni algoritem za nize oznak, ki je opisan v RFC 4647.

Tako bo poiskal naslednjo najboljšo rešitev in skrajšal *sl-rozaj* na *sl*, kar bi privedlo do HTML-ja v slovenščini. Rezijski uporabnik se bo čudil, zakaj je dobil toliko strani v slovenščini, za katere ni zaprosil, dobil pa bo zelo malo italijanskih strani, čeprav je določil italijanščino kot drugo možnost. Spomnimo se, da je ravno to ena izmed točk definicije jezika v ISO 639-3: »Dve podobni različici sta po navadi ocenjeni za različici istega jezika, če so govorniki vsake različice sposobni razumevanja druge različice (tj. če lahko na osnovi znanja svoje različice razumejo drugo različico, ne da bi se je morali učiti) na funkcionalni ravni.« V primeru rezijske in slovenščine to inherentno razumevanje ni dano, kot so se strinjali tudi vsi prisotni na srečanju na SAZU v Ljubljani decembra 2006.

6 Politične razsežnosti standardov informacijske tehnologije

Zunaj okolja strokovnih diskusij je bil predmet kodne sestavine ISO za rezijsko izredno politično vnetljiva tema:

- (1) Zdi se, da obstoj registrirane kode ISO 639 za ime jezikovne različice izboljša sam status različice. Tako kot se širši javnosti zdi, da imeti ali ne imeti slovar ali slovnico pomeni razliko med jezikom in samo narečjem, se ji tudi zdi, da obstoj ali neobstoj kodnega elementa za jezikovno različico povzroča enako hierarhično razlikovanje. To je tudi razumljivo, saj je širša javnost obveščena le o kodnih sestavinah pogosto rabljenih knjižnih jezikov. Da pa je Slovenska akademija znanosti in umetnosti padla pred istim predsodkom, je obžalovanja vredno.
- (2) V nasprotju z avtorjevo namero je osnutek prišel v širšo javnost, zato se ni bilo mogoče izogniti zgoraj navedenim napačnim interpretacijam. Avtor pa dvomi, da so bile teme, kot so podatkovni zapis, pravilni nizi oznak, medsebojno razumevanje in njegove posledice za ujemanje nizov oznak sploh predmet širše diskusije.
- (3) Kodni sestavina za rezijsko in v okviru ISO 639-3 bi tudi z znanstvenega vidika potrdila obstoj očitnih razlik med slovenščino in rezijsko. Prav te razlike rezijsko nekoliko ločujejo od drugih pokrajinskih pisnih oblik slovenščine zunaj Republike Slovenije, kot sta nadiščina in porabščina. Za govorce teh dveh jezikovnih različic sporazumevanje z drugimi Slovenci ni taka težava. Vendar pa to, da rezijska nima registrirane kodne sestavine, ne odpravi obstoječih razlik. Zanihanje obstoja razlik še ne pomeni, da razlik res ni.
- (4) Kodne sestavine ISO 639 so predstavniki imen jezikovnih različic, odsotnost kodne sestavine pa pomeni, da ni možno poimenovati ustrezne jezikovne različice. Nihče ne zanika obstoja pojma *rezijska*, torej bi morala zaradi učinkovite komunikacije obstajati tudi možnost, da se ga označi s formalno sestavino, kot je samostalnik ali pa kodna sestavina. Ali naj nekaterih znanstvenih tem ne bi načenjali zaradi političnega naboja, ki ga lahko imajo za neobveščeno ali slabo obveščeno javnost? Ali pa bibliografska zbirka podatkov iz istega razloga naj ne bi bila sposobna vrniti dokumentov?

(5) Nevložitev prošnje za kodni element za rezijanščino pri ISO 639-3 je učinkovito preprečila, da bi lahko mednarodna skupnost kaj prispevala k diskusiji o vprašanju rezijanščine v razmerju do slovenščine. Se bojimo mednarodnega mnenja ali pa nam je prav malo mar, kaj si svet misli o slovenščini?

7 Pogled naprej

Da bi vendar končali v pozitivnem vzdušju, naj nakažemo nekaj možnosti za nadaljnje ravnanje. Po avtorjevem mnenju bi bilo treba začeti s širšo diskusijo v zvezi z različnimi pristopi ISO 639-3, RFC 5646 in v prihodnosti tudi ISO 639-6 o enkodiranju podstandardnih jezikovnih različic, da bi se našel najprimernejši, ki bi najbolje služil potrebam slovenskih jezikoslovcev in odseval slovenski jezikovni položaj. Tej diskusiji naj bi se pridružili tudi strokovnjaki za informacijske tehnologije. Tule je nekaj okvirnih idej.

(1) Za vsako kodno sestavino ISO 639-3 in IANA Language Subtag Registry je treba zaprositi. Prosilcu pa je prepuščeno, ali bo izbral izbirni ali popolni pristop. Praksa kaže, da so prosilci raje selektivni in zadostijo le svojim trenutnim potrebam. ISO 639-6 ponuja izčrpen seznam kodnih sestavin za vse jezikovne različice, vključno z narečji. Pričakovati pa je, da ta kodni seznam ne bo ostal nespremenjen, saj se sčasoma tudi znanje o narečjih širi.

(2) ISO 639-6 je drugačen po obsegu kodnih sestavin, saj se lahko ena štiricrkovna koda hkrati nanaša na atribute kot »hierarhični položaj«, »govorjeni ali pisni jezik«, »pisava«, »zgodovinsko stanje«. Ta sinkretizem je v velikem kontrastu do analitičnega pristopa RFC 5646, ki se raje nanaša na te atribute s sredstvi razširjenih jezikovnih oznak in variantnih podoznak (oba za hierarhični položaj) in oznake za pisavo. Na zgodovinsko stopnjo se včasih nanaša koda ISO 629-2 ali 639-3, kot je *goh* 'stara visoka nemščina' (< *Old High German*) in *gmh* 'srednja visoka nemščina' (< *Middle High German*), včasih pa se nanj nanaša variantna podoznaka.

(3) Kodne sestavine ISO 639-6 imajo dolžino niza določeno, v primerjavi s pravnimi nizi oznak RFC 5646, ki so lahko različno dolgi in ki lahko postanejo zelo dolgi. Zato lahko v določenih okoljih pride do okrnitve takega niza oznak.

(4) Hierarhije, ki jih predstavlja ISO 639-6, imajo nekaj inherentnih problemov, saj pokrivajo različne parametre jezikovne variantnosti: zemljepisne, časovne in sporočanje. Zemljepisno narečje ima govorno različico jezika za svojega prednika, kot je »govorjena angloangleščina« za »scouse«; kaj pa, če je narečje uporabljeno tudi v pisni obliki? Podobno imajo zgodovinske jezikovne različice konsistentno kot svojega prednika pisno različico, kar povzroča nemalo težav pri umeščanju zapisov zgodovinskih narečij v hierarhijo.

(5) Trenutno še ni popolnoma jasno, kako naj bi bili oblikovani pravilni nizi oznak s kodnimi sestavinami ISO 639-6. To resno omejuje obseg praktične rabe teh kod.

Sicer pa bi bilo treba razmišljati o razlikovalnih značilnostih na območju slovenskega jezika, na katere bi se nanašale registrirane kodne sestavine. Najpomembnejše so gotovo »narečna skupina«, »narečje«, »krajevni govor«, »pisava«, »pravopis«, »zgodovinsko stanje«, »uradni ali pokrajinski pisni jezik«, zamislimo pa si lahko tudi druge vrste razlikovanj.

Sledi hipotetična rešitev za prva štiri razlikovanja, če izhajamo iz ISO 639-3 v kombinaciji z RFC 5646, ki predstavljata trenutno stanje področja. Po avtorjevem mnenju v tem trenutku ISO 639-6 ni dobra možnost, saj ostaja nerešenih nekaj inherentnih problemov in ni jasno, kako bi se vključil v RFC 5646. Vse kodne sestavine v spodnjih nizih oznak so rešitev ad hoc in služijo le kot primer. Edini izjemi sta *sl* in *njiva*.

Ker so v slovenščini razlike med narečji precej velike, bi morda moral biti prvi korak, da se interpretira koda ISO 639-3 *sl/slv* kot koda, ki se nanaša na makrojezik, prav tako kot že *sqi* za albanščino in *srd* za sardščino. Narečne skupine in izolirana narečja, kot je rezijanščina, imajo lahko dodeljeno kodno sestavino ISO 639-3, ki se jo lahko registrira kot razširjeno jezikovno oznako pod RFC 5646. Koroško narečno skupino bi se zdaj dalo kodirati kot *sl-kor* ali le *kor*, rezijanščino pa kot *sl-roz* ali le *roz*. Narečja, ki nesporno pripadajo narečni skupini, in lokalna narečja bi lahko enkodirali z variantno podoznako pod RFC 5646 kot *sl-kor-ziljs* ali *kor-ziljs* za 'ziljsko narečje' in *sl-kor-ziljs-ovcja* ali *kor-ziljs-ovcja* za krajevni govor Ovcje vasi. Možno bi bilo neposredno zaporedje makrojezikove oznake *sl/slv* in variantne podoznake, tj. *sl-ziljs*, tako da bi izpustili razširjeno jezikovno oznako in možno sledečo variantno podoznako.

Pisava je vedno latinica, saj se pisavne različice, ki so bile uporabljene za slovenščino, razlikujejo v vključevanju specifičnih črk in ne v splošnih oblikah. Ker je *Latn* torej za slovenščino odvečna oznaka, je nepotrebna in ne sme biti vključena v niz oznak. Za različne pisave v slovenščini se lahko uporabijo različne podoznake, kot so *bohori*, *metelc*, *dajnci* in *gajica*. V zgoraj navedenih primerih pa ne more biti specificirana velika začetnica, kot za oznake za pisavo priporoča specifikacija RFC 5646 (*Latn*, *Latf* itd.). Da bi bile variantne podoznake lažje prepoznavne, bi slovenske pisave morale imeti neke druge formalne značilnosti, kot je določena dolžina ali pa določeno zaporedje črk. Tako bi se lahko tvorile oznake kot *sl-metelc* za zgodovinsko pisavo slovenščine 19. stoletja ali *roz-bohori-njiva* za najstarejše poznano besedilo, napisano v rezijanščini. Vsekakor pa bi bilo treba soglašati o nekem notranjem urejanju tipov variantnih podoznak. V navedenem primeru je imitiran položaj za oznake RFC 5646 za pisave: stoji tik za možno razširjeno jezikovno oznako in pred vsemi drugimi podoznakami.

Če se slovenski znanstveni svet ne bo prebudil, obstaja možnost, da uradna standardizacija v ISO 639, ki zadeva slovenščino, ne bo ustrezala njegovim potrebam. Dovolj je pogledati podatkovni zapis o slovenščini v ISO 639-3 [ISO 639-3 *slv*], kjer je takoj jasno, da je za izboljšave še veliko prostora.

Prevedla Polona Liberšar

Viri

- DCMI: <http://dublincore.org/>
ES = *Enciklopedija Slovenije* 1–15, Ljubljana: Mladinska knjiga, 1987–2001.
IANA Language Subtag Registry = <http://www.iana.org/assignments/language-subtag-registry>
IETF Language tag discussions = <http://www.alvestrand.no/mailman/listinfo/ietf-languages>
ISO 639-1 = http://www.infoterm.info/standardization/iso_639_1_2002.php
ISO 639-2 = <http://www.loc.gov/standards/iso639-2/>
ISO 639-3 = <http://www.sil.org/iso639-3/>
ISO 639-3 slv = http://www.ethnologue.com/show_language.asp?code=slv
ISO 639-6 = <http://www.geolang.com/aspbite/categories/index.asp?intCatID=19>
ISO 3166-1 = http://www.iso.org/iso/country_codes.htm
ISO 3166-2 = http://www.iso.org/iso/country_codes/background_on_iso_3166/iso_3166-2.htm
ISO 15924 = <http://unicode.org/iso15924/index.html>
ISO FDIS 639-6 en-GB = <http://www.ietf.org/mail-archive/web/ltru/current/msg13138.html>
MS 1 = ISO639-3_ChangeRequest_SLV.doc (priloga 1; zadnja sprememba 6. decembra 2006)
MS 2 = ISO639-3_NewCodeRequest_Resian.doc (priloga 2; zadnja sprememba 6. decembra 2006)
Resianica = <http://purl.org/resianica/>
RFC 3066 = <http://www.ietf.org/rfc/rfc3066.txt>
RFC 4646 = <http://www.ietf.org/rfc/rfc4646.txt>
RFC 4647 = <http://www.ietf.org/rfc/rfc4647.txt>
RFC 5646 = <http://www.rfc-editor.org/rfc/rfc5646.txt>
SSKJ = *Slovar slovenskega knjižnega jezika* 1–5, Ljubljana: SAZU oz. ZRC SAZU (izd.) – DZS (zal.), 1970–1991.

The meaning and scope of tags for linguistic varieties in information technology

Summary

Common nouns and proper nouns refer to concepts that single out concrete and abstract objects in the real world. For example, the proper nouns *Resian* and *Slovenian* refer to linguistic varieties encountered in the real world. The hierarchic relations between these objects are not an intrinsic quality of the proper nouns themselves, but of the concepts *rezijanščina* and *slovenščina* that the proper nouns refer to. The same holds for the abbreviations for these proper nouns.

ISO 639, divided up into various substandards, registers abbreviations for linguistic varieties called language codes. ISO 639-1 (for widely used standard languages) and 639-2 (for less used standard and non-standard languages) have no clear criteria for distinguishing between languages and dialects and lack explicit records with information on the individual language codes: the concepts to which the codes refer remain outside the scope of these substandards. ISO 639-1 and 639-2 do not “register languages,” but only codes for language names. On the other hand, ISO 639-3 aims to produce and maintain a comprehensive list of all known human languages, spoken and written. Every code element has an information record associated with it. Thus, both code and concept are an intrinsic part of this substandard. Furthermore, it has an explicit approach to the question of language vs. dialect. Finally, ISO 639-6 is oriented toward a fine-grained distinction of as many linguistic varieties as possible, be these of a dialectal, historical, or graphic nature.

RFC 5646 and RFC 4647 have been formulated in order to use the language codes registered with ISO 639-1, 639-2, and 639-3 within tag strings for describing information objects.

Although the proposal for applying for a language code for Resian has recently met with fierce resistance, it is still thought that the framework presented above can be used to create and register language codes for the linguistic varieties that can and should be distinguished within Slovene.

Priloga 1 = MS1

ISO 639-3 Registration Authority: Request for Change to ISO 639-3 Language Code Change Request Number: (completed by Registration authority)

Date: 2006-12-31

Primary Person submitting request: Han Steenwijk

Affiliation: University of Padova

E-mail address: han.steenwijk@unipd.it

Names, affiliations and email addresses of additional supporters of this request:

Roberto Dapit, University of Udine, roberto.dapit@uniud.it

Aleksandr D. Duličenko, University of Tartu, aleksd@list.ru

Matej Šekli, University of Ljubljana, matej.sekli@guest.arnes.si

Jože Toporišič, Slovenian Academy of Arts and Sciences, darja.pate@sazu.si

Postal address for primary contact person for this request (in general, email correspondence will be used):

prof. Han Steenwijk

Universita di Padova

Dipartimento di Lingue e Letterature Anglo-Germaniche e Slave

Sezione di Slavistica

Via Beldomandi, 1

I-35139 Padova

Italy

Do not be concerned about your responses causing the form text spacing or pagination to change. Use Shift-Enter to insert a new line in a form field (where allowed)

Types of change requests

This form is to be used in requesting changes (whether creation, modification, or deletion) to elements of the ISO 639 Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages. The types of changes that are possible are to 1) modify the reference information for an existing code element, 2) retire a code element from use, including merging its scope of denotation into that of another code element, 3) split an existing code element into two or more separate language code elements, or 4) create a new code element. Fill out section 1, 2, 3, or 4 below as appropriate, plus section 5 to document the sources of your information. The process by which a change is received, reviewed and adopted is summarized on the final page of this form.

For changes of type 1, 2 or 3 (proposing a change to an existing code), please identify:

Affected ISO 639-3 identifier: SLV
Associated reference name: Slovenian

1. Modify an existing language code element

- (a) What are you proposing to change:
- Language name
 - Language type (living, extinct, historical, etc.)
 - Language scope (individual language or macrolanguage)
- (b) What new value(s) do you propose:
- (c) Rationale for change:

2. Retire a language code element from use

- (a) Reason for change:
- There is no evidence that the language exists.
 - This is equivalent to another ISO 639-3 language.
 - This is a variety that is fully intelligible with another ISO 639-3 language and should be merged with it.
- (b) If one of the latter two reasons, with which ISO 639-3 code element (identifier and name) should it be merged :
- (c) Rationale for change:

3. Split a language code element into two or more code elements

- (a) List the languages into which this code element should be split:
Slovenian, Resian

By the language identification criteria set forth in ISO 639-3, the simple fact of distinct identities is not enough to assign separate identifiers. The criteria are defined in the standard as follows:

For this part of ISO 639, judgments regarding when two varieties are considered to be the same or different languages are based on a number of factors, including linguistic similarity, intelligibility, a common literature (traditional or written), a common writing system, the views of users concerning the relationship between language and identity, and other factors. The following basic criteria are followed:

- *Two related varieties are normally considered varieties of the same language if users of each variety have inherent understanding of the other variety (that is, can understand based on knowledge of their own variety without needing to learn the other variety) at a functional level.*
- *Where intelligibility between varieties is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variety*

that both understand can be strong indicators that they should nevertheless be considered varieties of the same language.

- *Where there is enough intelligibility between varieties to enable communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages*

(b) Referring to the criteria given above, give the rationale for splitting the existing code element into two or more languages:

1. Literary Slovenian and Resian are not mutually intelligible without prior learning, but with speakers of neighbouring Slovenian dialects few problems of intelligibility exist. A speaker with Slovenian as his first language will find Resian very difficult to understand, while a speaker with Resian as his first language will understand some basic Slovenian.
2. Literature printed in Slovenian is only understood by a small number of speakers of Resian and is not regarded by them as part of their ethnolinguistic identity.
3. Although speakers of Slovenian consider Resian to be a Slovenian dialect, most Resians do not share this point of view.

In order to complete the change request, the form “Request for New Language Code Element in ISO 639-3” (file name “ISO639-3_NewCodeRequestForm.doc” or “ISO639-3_NewCodeRequestForm.rtf”) must also be submitted for each new identifier that is to be created. That step can be deferred until this form has been processed by the ISO 639-3 registrar.

4. Create a new language code element

- (a) Name of missing language:*
- (b) State the case that this language is not the same as any language that already has an identifier in ISO 639-3:*

In order to complete the change request, the form “Request for New Language Code Element in ISO 639-3” (file name “ISO639-3_NewCodeRequest.doc” or “ISO639-3_NewCodeRequestForm.rtf”) must also be submitted to more fully document the new language.

5. Sources of information

Please use whichever of the points below are relevant in order to document the sources on which you have based the above proposal.

- (a) First-hand knowledge. Describe:*
The requester is doing fieldwork on Resian since 1987, returning to the Resia Valley every year from then on.

- (b) *Knowledge through personal communication. Describe:*
(c) *Knowledge from published sources (please give complete bibliographical references):*

- Roberto Dapit. "Nastajanje krajevnih knjižnih jezikov pri Slovencih v Furlaniji", in: Slovenski knjižni jezik: aktualna vprašanja in zgodovinske izkušnje (= Obdobja 20), uredila A. Vidovič Muha, Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, 2003, pp. 301-312.
- Aleksandr D. Duličenko. "Das Resianische", in: Einführung in die slavischen Sprachen, 3. Auflage, herausgegeben von P. Rehder, Darmstadt: Wissenschaftliche Buchgesellschaft, 1998, pp. 246-249.
- Tine Logar. "Solbica (Stolizza; OLA 1)", in: Fonološki opisi srpskohrvatskih/hrvatskosrpskih, slovenačkih i makedonskih govora obuhvaćenih opšte-slovenskim lingvističkim atlasom, Sarajevo, 1981, pp. 35-40.
- Pavle Merku. "Enkratnost Rezije: Navezanost na kulturne korenine in shizoglosija", Jezik in slovstvo 31/4, 1985/86, pp. 101-107.
- Fran Ramovš. "The Resianic dialect", in: The National Frontier Between Slovenia and Italy: Linguistic and Social Features, Ljubljana: J. Blaznika nasl. d. d., 1945, pp. 5-7.
- Matej Šekli. "Jezik, knjižni jezik, pokrajinski oz. krajevni knjižni jezik: genetskojezikoslovni in družbenostnojezikoslovni pristop k členjenju jezikovne stvarnosti (na primeru slovenščine)", in: Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti (= Obdobja 22), uredila E. Kržišnik, Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, 2004, pp. 41-58.
- Han Steenwijk. "Resian as a minority language", in Language Death and Language Maintenance: Theoretical, practical and descriptive approaches (= Current Issues in Linguistic Theory, vol. 240), edited by M. Janse & S. Tol, Amsterdam - Philadelphia, John Benjamins, 2003, pp. 215-226.
- Zuljan, Irena. "Rezija in rezijanščina sta res enkratni: sociolingvistični izsledki jezikoslovne skupine mladinskega raziskovalnega tabora Rezija 1989", Primorska srečanja 102, 1989, pp. 913-918.
- Zybatow, Lew N. "Resianisch: eine bedrohte Mikrosprache im Herzen Europas", in: Neue Forschungsarbeiten zur Kontaktlinguistik (= Plurilingua 19), herausgegeben von W.W. Moelleken & P.J. Weber, Bonn: Dümmler, 1997, pp. 505-514.

The change proposal process

A request to change the code set goes through a five step process:

- 1. A user of ISO 639-3 proposes a change and submits it to the ISO 639-3 Registration Authority (ISO 639-3/RA) using this form.*
- 2. The ISO 639-3 registrar processes the change request to verify that the request is compatible with the criteria set forth in the standard and to ensure that the submit-*

ter has supplied all necessary information. This may involve rounds of interaction with the submitter.

3. When the change request proposal is deemed to be sufficiently complete in its documentation (including all associated New Code Requests), the change request is promoted to "Proposed Change" status and the ISO 639-3 registrar posts the request on the official web site of the ISO 639-3/RA. Also at this time, an announcement is sent to appropriate discussion lists (those which express in their statement of purpose an interest in the language or language family in focus in the proposal, inviting individuals to review and comment on the proposal. The source of list information is the LinguistList Mailing List index, <http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/lists/index.html>. Responses are sent to the ISO 639-3 registrar for compilation. The consensus of reviews may result in promotion to "Candidate Status" (with or without amendment), or withdrawal of the change request, if the conclusion is that the request is not in keeping with the stated criteria of the ISO 639-3 standard.

4. Three months prior to the end of the annual cycle of review and update, a new notice is posted on the official web site of the ISO 639-3/RA, and an announcement listing the Candidate Status Change Requests is posted to LINGUIST, <http://linguistlist.org/issues/index.html>, and other general language and linguistics discussion lists, as well as lists relating to specific languages or language families. All change requests are then open to further review and comment by any interested party for a period of three months.

5. At the end of the formal review period, a given Change Request may be: 1) adopted as a whole, 2) adopted in part (specific changes implicit in the whole Change Request may be adopted separately), 3) amended and resubmitted for the next review cycle, or 4) withdrawn from consideration. All change requests remain permanently archived at the official web site of the ISO 639-3/RA.

Please return this form to:

ISO 639-3 Registrar

SIL International, Office of Language Information Systems

7500 West Camp Wisdom Road

Dallas, Texas 75236 USA

ISO 639-3/RA web site: <http://www.sil.org/iso639-3/>

E-mail: iso639-3@sil.org

An email attachment of this completed form is preferred.

Sources of documentation for ISO 639-3 identifiers:

Gordon, Raymond G., Jr. (ed.), 2005. *Ethnologue: Languages of the World*, Fifteenth edition. Dallas, Tex.: SIL International. Online version: <http://www.ethnologue.com/>.

LinguistList. *Ancient and Extinct Languages*. <http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/forms/langs/GetListOfAncientLgs.html>

LinguistList. *Constructed Languages*. <http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/forms/langs/GetListOfConstructedLgs.html>

Priloga 2 = MS2

ISO 639-3 Registration Authority:

Request for New Language Code Element in ISO 639-3

Date: 2006-12-31

Name of Primary Requester: Han Steenwijk

E-mail address: han.steenwijk@unipd.it

Names, affiliations and email addresses of additional supporters of this request:

Roberto Dapit, University of Udine, roberto.dapit@uniud.it

Aleksandr D. Duličenko, University of Tartu, aleksd@list.ru

Matej Šekli, University of Ljubljana, matej.sekli@guest.arnes.si

Jože Toporišič, Slovenian Academy of Arts and Sciences, darja.pate@sazu.si

Associated Change request number : (completed by Registration Authority)

Tentative assignment of new identifier : (completed by Registration Authority)

Do not be concerned about your responses causing the form text spacing or pagination to change. Use Shift-Enter to insert a new line in a form field (where allowed)

1. NAMES and IDENTIFICATION

- (a) *Preferred name of language for code element denotation:*
Resian
- (b) *Autonym (self-name) for this language:*
rozajanski
- (c) *Alternate names and spellings of language, and any established abbreviations:*
rozojanski; rosajanski, rosojanski
- (d) *Reason for preferred name:*
This name is commonly used in English literature on the subject.
- (e) *Name of ethnic group or description of people who use this language and approximate population of users, if in use today:*
Resians (It. Resiani, Sl. rezijani) living in the Resia Valley, Udine Province, Italy and as emigrants throughout the world; plm. 2000 speakers
- (f) *Preferred three letter identifier, if available:* roj

Your suggestion will be taken into account, but the Registration Authority will determine the identifier to be proposed. Language identifiers are not intended to be an abbreviation for a name of the language, but to serve as a device to identify a given language uniquely. With thousands of languages, many sets of which have similar names, it is not possible to provide identifiers that resemble a language name in every case.

2. TEMPORAL DESCRIPTION and LOCATION

- (a) *Is this a* *Living language*
 Nearly extinct/secondary use only (includes languages in revival)
 Recently extinct language
 Historical language
 Ancient language
 Artificially constructed language

(*Select one. See explanations of these types at <http://www.sil.org/iso639%2D3/types.asp>*)

- (b) *Countries where used:*
Italy
- (c) *Region within each country: towns, districts, states or provinces where used:*
Comune di Resia, Provincia di Udine, Regione Autonoma Friuli-Venezia Giulia
- (d) *For an ancient or historical language, give approximate time frame; for a recently extinct language, give the approximate date of the last known user's death*

3. MODALITY AND LINGUISTIC AFFILIATION

- (a) *This language is:*
 Signed *Spoken* *Attested only in writings*
- (b) *Language family, if classified; origin, if artificially constructed:*
Indo-European, Slavic, South Slavic, Western South Slavic, Slovenian, North-western Slovenian
- (c) *Closest language linguistically. For signed language, note influence from other signed or spoken languages:*
North-western Slovenian dialects

4. LANGUAGE DEVELOPMENT AND USE

- (a) *What written literature, inscriptions or recordings exist in this language? Are there newspapers, radio or television broadcasts, etc.?:*
Resian written texts have been produced from the 18th century onwards. Up until the 1970's, these texts were of mainly religious character.
Two Resian periodicals, "All'ombra del Canin" and "Náš glas" regularly contain texts written in the local language.
The weekly radio programme in Resian last 30 minutes and is broadcasted by RAI on Saturdays.

- (b) *Is this language officially recognized by any level of government? Is it used in any levels of formal education as a language of instruction (for other subjects)? Is it taught in schools?:*
Resian is not officially recognised and is not used as a language of instruction. It is, however, taught in elementary school as a subject.
- (c) *Comment on factors of ethnolinguistic identity and informal domains of use:*

5. SOURCES OF INFORMATION

You do not need to repeat sources previously identified in the form, "Request for Change to ISO 639-3 Language Code"

- (a) *First-hand knowledge. Describe:*
(b) *Knowledge through personal communication. Describe:*
(c) *Knowledge from published sources. Include known dictionaries, grammars, etc. (please give complete bibliographical references):*

Jan Baudouin de Courtenay. Opyt fonetiki rez'janskich govorov. Varšava-Peterburg: Vende i Ko. - Kožančikov, 1875.

Sergio Chinese. Rošajanskë - laškë bysidnjäk/Repertorio lessicale italiano - resiano. Udine: Marioni, 2003.

Tine Logar. "Rezijanski dialekt: glasoslovna skica", in: VIII. seminar slovenskega jezika, literature in kulture", Ljubljana, 1972. pp. 1-10.

Fran Ramovš. "Karakteristika slovenskega narečja v Reziji". Časopis za slovenski jezik, književnost in zgodovino 7, 1928. pp. 107-121.

Han Steenwijk. The Slovene Dialect of Resia: San Giorgio (= Studies in Slavic and General Linguistics Volume 18), Amsterdam - Atlanta, GA: Rodopi, 1992.

Han Steenwijk. Ortografia resiana/Tö još rozajanskë pisanjë, Padova, CLEUP 1994.

Han Steenwijk. Grammatica pratica resiana: il sostantivo, Padova, CLEUP 1999.

Han Steenwijk. Piccolo dizionario ortografico resiano/Mali bisidnik za tö još rozajanskë pisanjë, Padova, CLEUP 2005.

Please return this form to:

ISO 639-3 Registrar

SIL International, Office of Language Information Systems

7500 West Camp Wisdom Road

Dallas, Texas 75236 USA

ISO 639-3/RA web site: <http://www.sil.org/iso639%2D3/default.asp>

Email: iso639-3@sil.org

An email attachment of this completed form is preferred.

Further information:

If your request for a new language code element is supported by the Registration Authority as a formal proposal, you may be contacted separately by researchers working with the Ethnologue or with LinguistList asking you to provide additional information.

Sources of documentation for ISO 639-3 identifiers:

Gordon, Raymond G., Jr. (ed.), 2005. Ethnologue: Languages of the World, Fifteenth edition. Dallas, Tex.: SIL International. Online version: <http://www.ethnologue.com/>.

LinguistList. Ancient and Extinct Languages. <http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/forms/langs/GLinguistList>.

Constructed Languages. <http://cf.linguistlist.org/cfdocs/new-website/LL-WorkingDirs/forms/langs/GetListOfConstructedLgs.html>